

This is a repository copy of *Comparing predictive accuracy in small samples using fixed-smoothing asymptotic*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/155897/>

Version: Published Version

---

**Article:**

Coroneo, Laura [orcid.org/0000-0001-5740-9315](https://orcid.org/0000-0001-5740-9315) and Iacone, Fabrizio [orcid.org/0000-0002-2681-9036](https://orcid.org/0000-0002-2681-9036) (2020) Comparing predictive accuracy in small samples using fixed-smoothing asymptotic. *Journal of Applied Econometrics*. pp. 391-405. ISSN 0883-7252

<https://doi.org/10.1002/jae.2756>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

RESEARCH ARTICLE

# Comparing predictive accuracy in small samples using fixed-smoothing asymptotics

Laura Coroneo<sup>1</sup> | Fabrizio Iacone<sup>2,1</sup>

<sup>1</sup>Department of Economics and Related Studies, University of York, York, UK

<sup>2</sup>Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano, Milan, Italy

## Correspondence

Laura Coroneo, Department of Economics and Related Studies, University of York, Heslington, York YO10 5DD, UK.  
Email: laura.coroneo@york.ac.uk

## Summary

We consider fixed-smoothing asymptotics for the Diebold and Mariano (*Journal of Business and Economic Statistics*, 1995, 13(3), 253–263) test of predictive accuracy. We show that this approach delivers predictive accuracy tests that are correctly sized even when only a small number of out-of-sample observations is available. We apply the fixed-smoothing asymptotics to the Diebold and Mariano test to evaluate the predictive accuracy of the Survey of Professional Forecasters (SPF) and of the European Central Bank Survey of Professional Forecasters (ECB SPF) against a simple random walk. Our results show that the predictive abilities of the SPF and of the ECB SPF were partially spurious.

## 1 | INTRODUCTION

Good forecasts are key to good decision making, and being able to compare predictive accuracy is key to discriminating between good and bad forecasts. To this end, one of the most used tests to compare the predictive accuracy of two competing forecasts is the Diebold and Mariano (1995; DM) test.

The DM test is based on a loss function associated with the forecast errors of each forecast, testing the null hypothesis of zero expected loss differential between two competing forecasts. This framework allows us to test for equal predictive accuracy using any loss function, and the test statistic is valid for contemporaneously correlated, serially correlated, and nonnormal forecast errors.

The DM approach takes forecast errors as model free, and the test is valid also when the forecasts are produced from unknown models, as for example from forecast survey data. When the forecasts are produced by estimated models, nested or nonnested, it is in general necessary to account for the impact of the model parameter estimation uncertainty on the distribution of the forecast accuracy test (see Clark & McCracken, 2001; West, 1996). In this case, the limiting distribution of the test statistics depends on the specific modeling assumptions made for obtaining the forecast errors (see Clark & McCracken, 2013; West, 2006). West (1996) showed that in some cases the DM approach was asymptotically valid even when forecasts were obtained from estimated models. This happens when the number of in-sample observations is large relative to the number of out-of-sample observations or when in-sample and out-of-sample loss is the same, for example using a quadratic loss as an evaluation criterion for models that are estimated by ordinary least squares (OLS). However, in practice, it is not uncommon to compare forecasts produced by models for which accounting for the model parameter estimation uncertainty is not tractable. In addition, if the objective is to compare forecasting methods as opposed to forecasting models, then Giacomini and White (2006) showed that in an environment with asymptotically nonvanishing estimation uncertainty the DM test can still be applied. For these reasons, the DM test is still widely applied also when forecasts are obtained by estimated models (see Diebold, 2015).

One additional reason for the success of the DM test is that the test statistic is simple to compute and asymptotically normally distributed. However, as also noted by DM, the test can be subject to large size distortions in small samples,

which can be spuriously interpreted as superior predictive ability for one forecast. This is due to the fact that in the test statistic the long-run variance is replaced by a consistent estimate, and standard limit normality is then still employed, this may be unsatisfactory when only a small number of out-of-sample observations are available. As remarked by Clark and McCracken (2013), “one unresolved challenge in forecast test inference is achieving accurately sized tests applied at multistep horizons—a challenge that increases as the forecast horizon grows and the size of the forecast sample declines”.

In this paper, maintaining the framework of asymptotically nonvanishing estimation uncertainty of Giacomini and White (2006), we show theoretically, and we verify in a Monte Carlo exercise based on Giacomini and Rossi (2010), that this challenge can be resolved by using alternative asymptotics for the limit distribution of the scaled expected loss differential between two competing forecasts. The first alternative asymptotics is the fixed- $b$  approach of Kiefer and Vogelsang (2005), in which the limit properties of the weighted autocovariances estimate of the long-run variance are derived assuming that the bandwidth-to-sample size ratio is constant. With this approach, the test to compare predictive accuracy has a nonstandard limit distribution that depends on the bandwidth-to-sample ratio  $b$  and on the kernel used to estimate the long-run variance. The second alternative asymptotics that we consider is the fixed- $m$  approach as in Sun (2013) and Hualde and Iacone (2015, 2017). In this case, the estimate of the long-run variance is based on a weighted periodogram estimate with Daniell kernel and a truncation parameter  $m$  that is assumed to be constant as the sample size increases. The test to compare predictive accuracy has a  $t$  distribution with degrees of freedom that depend on the truncation parameter. This averaged periodogram estimate can be seen as one application of the orthonormal series variance estimate (see Phillips, 2005). Following Sun, (2014a, 2014b), we refer to these two alternative asymptotics, fixed- $b$  and fixed- $m$ , as “fixed-smoothing asymptotics”. With this type of asymptotics, the assumption on the bandwidth parameter implies that the estimate of the long-run variance is not consistent. However, inference is more precise than with heteroskedasticity- and autocorrelation-consistent (HAC) standard asymptotics, and therefore it is often referred to as “heteroskedasticity-autocorrelation robust” (HAR; see, e.g., Lazarus et al., 2018).

Fixed-smoothing asymptotics is a valuable improvement on standard asymptotics, especially when the sample size is not too large, which is usually the case for forecast evaluation. Harvey et al. (2017) performed an extensive Monte Carlo simulation exercise to examine the small-sample size and power properties of the fixed- $m$  approach in this context. Their results indicate that the fixed- $m$  approach outperforms the modified statistic and critical value proposed by Harvey et al. (1997) to improve the small-sample size behavior of the test. In larger samples, Patton (2015) and Li and Patton (2018) showed that fixed- $b$  asymptotics delivered considerable size improvements for high-frequency forecast evaluations. Our paper is closer to Harvey et al. (2017), for the emphasis on small samples; moreover we consider fixed- $b$  asymptotics. We also show that fixed-smoothing asymptotics can be justified not only when forecasts are taken as primitive, but also in the more realistic case in which forecasts are generated by a model. We also allow for a much more general level of heterogeneity in the data-generating process (DGP), as in Giacomini and White (2006).

We perform a Monte Carlo study to analyze the size and power properties of the proposed tests of equal predictive accuracy in small samples. We revisit the size and power properties of the standard DM test in an environment with asymptotically nonvanishing estimation uncertainty, as in Giacomini and White (2006), using the Monte Carlo framework proposed by Giacomini and Rossi (2010). Our results indicate that: (i) fixed-smoothing asymptotics delivers correctly sized predictive accuracy tests for correlated loss differentials even in small samples; (ii) the power of the tests with fixed-smoothing asymptotics mimics the size-adjusted power and is comparable to the power of bootstrap tests. In summary, fixed-smoothing asymptotics, when coupled with an appropriate bandwidth choice, addresses the problem of relevant size distortions that is typically observed when standard asymptotics is used, especially in small samples and in the presence of serially correlated errors.

To illustrate the usefulness of fixed-smoothing asymptotics for equal predictive accuracy tests, we evaluate the predictive accuracy of the Surveys of Professional Forecasters (SPF) and the European Central Bank Survey of Professional Forecasters (ECB SPF) against a naive random walk. As for the SPF, we evaluate forecasts for four core macroeconomic indicators (output growth, inflation, the unemployment rate and the 3-month Treasury bill rate) for the period from 1987:Q1 until 2016:Q4. Results show that part of the superior predictive accuracy of the SPFs indicated by the DM test is spurious, especially in the most recent subsample. As for the ECB SPF, we evaluate forecasts for year-on-year euro area gross domestic product (GDP) growth and year-on-year euro area Harmonized Index of Consumer Prices (HICP) inflation for the period from 2006:Q1 to 2016:Q4. With such a small sample size, standard tests of equal predictive ability suffer from large size distortions and provide partially spurious results, which are not confirmed when using fixed-smoothing asymptotics.

The paper is organized as follows. In Section 2 we introduce the test for equal predictive accuracy, and in Section 3 we describe the DM estimate. In Section 4 we detail the tests for equal predictive accuracy using fixed- $b$  and fixed- $m$  asymptotics. In Section 5 we present our Monte Carlo study, including a Monte Carlo comparison with the bootstrap. In Section 6 we discuss the empirical applications, and in Section 7 we conclude. The Supporting Information Appendix

contains a wide Monte Carlo simulation, in which forecasts are taken as primitives, dedicated to the bandwidth selection problem.

## 2 | COMPARING PREDICTIVE ACCURACY

We consider the variable of interest  $y_t$ , for which we want to compare two  $h$ -step-ahead forecasts obtained from two alternative forecasting methods, based on some predictor variables  $x_t$ . We denote the observed vector by  $w_t \equiv (y_t, x_t)'$ , defined on a complete probability space  $(\Omega, \mathcal{F}, P)$ , and we denote the information set at time  $t$  by  $\mathcal{F}_t = \sigma(w'_1, \dots, w'_t)$ . The two  $h$ -step-ahead forecasts for time  $t$  are based on the information set  $\mathcal{F}_{t-h}$  and are denoted by  $\hat{y}_t^{(i)} \left( \hat{\delta}_{t-h, R_i}^{(i)} \right) \equiv f^{(i)}(w_{t-h}, w_{t-h-1}, \dots, w_{t-h-R_i+1}; \hat{\delta}_{t-h, R_i}^{(i)})$  for  $i = 1, 2$ , where the forecasts are measurable functions of a sample of size  $R_1$  for  $f^{(1)}$  and  $R_2$  for  $f^{(2)}$ . If a forecast is based on parametric models, the vector  $\hat{\delta}_{t-h, R_i}^{(i)}$  includes the estimates from the model. Otherwise,  $\hat{\delta}_{t-h, R_i}^{(i)}$  represents the semiparametric or nonparametric estimator used to construct the forecast.

We denote by  $L \left[ e_t^{(i)} \left( \hat{\delta}_{t-h, R_i}^{(i)} \right) \right]$ , for  $i = 1, 2$ , the loss associated with the forecast error  $e_t^{(i)} \left( \hat{\delta}_{t-h, R_i}^{(i)} \right) = y_t - \hat{y}_t^{(i)} \left( \hat{\delta}_{t-h, R_i}^{(i)} \right)$ ; for example, a quadratic loss would be  $L \left[ e_t^{(i)} \left( \hat{\delta}_{t-h, R_i}^{(i)} \right) \right] = \left[ e_t^{(i)} \left( \hat{\delta}_{t-h, R_i}^{(i)} \right) \right]^2$ . The null hypothesis of equal predictive ability of the two forecasting methods is then

$$H_0 : E \left\{ L \left[ e_t^{(1)} \left( \hat{\delta}_{t-h, R_1}^{(1)} \right) \right] - L \left[ e_t^{(2)} \left( \hat{\delta}_{t-h, R_2}^{(2)} \right) \right] \right\} = 0. \quad (1)$$

Let

$$d_t \equiv d_t \left( \hat{\delta}_{t-h, R_1}^{(1)}, \hat{\delta}_{t-h, R_2}^{(2)} \right) = L \left[ e_t^{(1)} \left( \hat{\delta}_{t-h, R_1}^{(1)} \right) \right] - L \left[ e_t^{(2)} \left( \hat{\delta}_{t-h, R_2}^{(2)} \right) \right]$$

denote the time- $t$  loss differential between the two forecasts and let

$$\bar{d} = \frac{1}{T} \sum_{t=R+h}^{R+h+T-1} d_t$$

denote the sample mean of the loss differential, where  $R \equiv \max(R_1, R_2)$ . We assume that the following assumptions hold.

**Assumption 1.**  $w_t$  is mixing with  $\phi$  of size  $-r/(2r-2)$ ,  $r \geq 2$ , or  $\alpha$  of size  $-r/(r-2)$ ,  $r > 2$ .

**Assumption 2.**  $R < \infty$ , as  $T \rightarrow \infty$ .

**Assumption 3.**  $E|d_t|^{2r} < \infty$  for all  $t$ .

**Assumption 4.**  $\sigma_T^2 \equiv \text{var}[\sqrt{T} \bar{d}] > 0$  for  $T$  sufficiently large.

*Remark 1.* Assumption 1 allows the data to be characterized by considerable heterogeneity and dependence. Assumption 2 implies the presence of asymptotically nonvanishing estimation uncertainty. Assumption 3 is standard in central limit theorems for mixing processes (see, e.g., Wooldridge & White, 1988), and it replaces the Lyapunov-type condition  $E|d_t|^{2+\delta} < \infty$ ,  $\delta > 0$ , for strictly stationary processes. Assumption 4 in this framework with asymptotically nonvanishing estimation uncertainty allows for both nested and nonnested models. This is the same framework considered by Giacomini and White (2006).

*Remark 2.* A particular case of the framework considered here is the one of DM, where forecasts are taken as primitive and do not depend on forecasting methods—as, for example, survey forecasts. In this case, Assumption 1 is understood to be for  $e_t^{(1)}$  and  $e_t^{(2)}$ .

DM and Giacomini and White (2006) showed that under  $H_0$  in Equation (1), if Assumptions 1-4 hold, then

$$\sqrt{T} \frac{\bar{d}}{\sigma_T} \rightarrow_d N(0, 1).$$

Unfortunately, this statistic is unfeasible in testing  $H_0$ , because  $\sigma_T^2$  is unknown. However, the parameter  $\sigma_T^2$  can be replaced with an appropriate estimate and, if a consistent estimate is used, then the limit normality is not affected by the replacement.

### 3 | THE DM ESTIMATE

A typical estimate for the long run variance is the weighted autocovariances estimate (WCE):

$$\hat{\sigma}^2 = \hat{\gamma}_0 + 2 \sum_{j=1}^{T-1} k(j/M) \hat{\gamma}_j, \quad (2)$$

where  $\hat{\gamma}_j = \frac{1}{T} \sum_{t=1}^{T-j} \hat{u}_t \hat{u}_{t+j}$ , with  $\hat{u}_t = d_t - \bar{d}$ , and  $k(\cdot)$  is a kernel function such that  $k(0) = 1$ ,  $|k(\tau)| < 1$ ,  $k(\tau) = k(-\tau)$ ,  $k(\tau)$  is continuous at  $\tau = 0$ , and  $\int_0^1 k(\tau)^2 d\tau < \infty$ . The parameter  $M$  is a bandwidth parameter (or a truncation lag), and for consistency of  $\hat{\sigma}^2$  regularity conditions include  $M \rightarrow \infty$  and  $M/T \rightarrow 0$  as  $T \rightarrow \infty$ . We refer to Hannan (1970) for a survey of these estimates, and for a discussion of which kernels ensure that  $\hat{\gamma}^2 \geq 0$ .

In a variation of this approach, DM noted that if  $\hat{y}_t^{(i)} \left( \hat{\delta}_{t-h, R_i}^{(i)} \right)$  was an optimal  $h$ -step-ahead forecast, then  $e_t^{(i)} \left( \hat{\delta}_{t-h, R_i}^{(i)} \right)$  was at most an  $\text{MA}(h-1)$ . Accordingly, they proposed setting  $M = h-1$  and  $k(j/M) = 1$  if  $j/M \leq 1$  and 0 otherwise, so

$$\hat{\gamma}_{\text{DM}}^2 = \hat{\gamma}_0 + 2 \sum_{j=1}^{h-1} \hat{\gamma}_j. \quad (3)$$

This does not meet the condition  $M \rightarrow \infty$ , but the estimate is nevertheless consistent, because it exploits the assumption that  $u_t \equiv d_t - E(d_t)$  is a stationary  $\text{MA}(h-1)$ , thus ensuring that under  $H_0$  in Equation (1)

$$\sqrt{T} \frac{\bar{d}}{\hat{\gamma}_{\text{DM}}} \rightarrow_d N(0, 1).$$

The choice of  $\hat{\gamma}_{\text{DM}}^2$  may be very appealing, as it exploits information about the structure of  $u_t$ . However, the rectangular kernel used in Equation (3) may generate negative estimates for  $\hat{\sigma}_{\text{DM}}^2$ , and this is undesirable. Moreover, the Monte Carlo exercise in DM suggests the possibility of large size distortions in small samples, which would be spuriously interpreted as evidence of superior predictive power for one forecast method. To avoid the risk of negative estimates of  $\sigma_T$ , DM mention the possibility of using alternative kernels and standard asymptotics, but simulations in Clark (1999), in which a Bartlett kernel was used, do not suggest that simply replacing the kernel results in a definite improvement of the size distortion.

## 4 | FIXED-SMOOTHING ASYMPTOTICS

### 4.1 | Fixed- $b$ asymptotics

Following the approach of Kiefer and Vogelsang (2005), we consider alternative asymptotics for the estimate in Equation (2): for given  $M$ , the ratio  $M/T$  is taken as fixed as  $T \rightarrow \infty$ . As  $M/T$  is fixed, letting  $b = M/T$ , this alternative approach is referred to as fixed- $b$  asymptotics. With this assumption, Kiefer and Vogelsang (2005) showed that the estimate of  $\sigma_T$  is not consistent and that the standardized sample mean has a nonstandard limit distribution that depends on  $b$  and on the kernel, they also provided a formula to generate quantiles of the limit distribution that can be used as critical values in tests.

For fixed- $b$  asymptotics and assuming that the Bartlett kernel is used, we introduce the following notation:

$$\hat{\gamma}_{\text{BART}}^2 = \hat{\gamma}_0 + 2 \sum_{j=1}^{T-1} k_{\text{BART}}(j/M) \hat{\gamma}_j, \quad M/T \rightarrow b, \quad (4)$$

$$k_{\text{BART}}(x) = \begin{cases} 1 - |x|, & \text{if } |x| \leq 1 \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

Kiefer and Vogelsang (2005) showed that

$$\text{if } b \in (0, 1], \text{ then } \sqrt{T} \frac{\bar{d}}{\hat{\gamma}_{\text{BART}}} \Rightarrow \Phi_{\text{BART}}(b), \quad (6)$$

where  $\Rightarrow$  denotes weak convergence in the  $D[0, 1]$  space with the Skorohod topology. They characterized the limit distribution  $\Phi_{\text{BART}}(b)$  and provided readily available critical values for various  $b$  and kernels. For the Bartlett kernel with  $b \leq 1$ , these can be obtained using the formula

$$q(b) = \alpha_0 + \alpha_1 b + \alpha_2 b^2 + \alpha_3 b^3,$$

where

$$\begin{aligned} \alpha_0 &= 1.6449, & \alpha_1 &= 2.1859, & \alpha_2 &= 0.3142, & \alpha_3 &= -0.3427 & \text{for } 0.950 \text{ quantile,} \\ \alpha_0 &= 1.9600, & \alpha_1 &= 2.9694, & \alpha_2 &= 0.4160, & \alpha_3 &= -0.5324 & \text{for } 0.975 \text{ quantile.} \end{aligned}$$

The results of Kiefer and Vogelsang (2005) provide asymptotics that may be valid for any  $M$ , even  $M = T$ , but note that Kiefer and Vogelsang do not automatically recommend using  $M = \lfloor bT \rfloor$ .<sup>1</sup> Rather, they provide alternative asymptotics for a user-chosen bandwidth. Thus, for example, assuming  $T = 128$  and  $M = \lfloor T^{1/3} \rfloor = 5$ , then  $b = 5/128 = 0.039063$  and the 5% critical value for a two-sided test is 2.0766 instead of 1.96.

When testing assumptions about the sample mean, Kiefer and Vogelsang (2005) showed in Monte Carlo simulations that the fixed- $b$  asymptotics yields a remarkable improvement in size. However, while the empirical size improves (it gets closer to the theoretical size) as  $b$  is closer to 1, the power of the test worsens, implying that there is a size–power tradeoff. These results are also confirmed analytically by Sun et al. (2008), who proved that the fixed- $b$  limit distribution provides a higher order correction.

## 4.2 | Fixed- $m$ asymptotics

We now consider an alternative estimate of the long run variance: a weighted periodogram estimate (WPE). Letting  $\lambda_j = 2\pi j/T$  for  $j = 0, \pm 1, \dots, \pm \lfloor T/2 \rfloor$  as the Fourier frequencies, and

$$I(\lambda_j) = \left| \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T d_t e^{-i\lambda_j t} \right|^2$$

as the periodogram of  $d_t$ , we consider estimates

$$\tilde{\sigma}^2 = 2\pi \sum_{j=1}^{\lfloor T/2 \rfloor} K_M(\lambda_j) I(\lambda_j), \quad (7)$$

where  $K_M(\lambda_j)$  is a kernel function that is symmetric and  $M$  is a bandwidth parameter.

Note that as  $\frac{1}{\sqrt{2\pi}} \sum_{t=1}^T \bar{d} e^{-i\lambda_j t} = \bar{d} \frac{1}{\sqrt{2\pi}} \sum_{t=1}^T e^{-i\lambda_j t}$  and, for  $j \neq 0$ ,  $\sum_{t=1}^T e^{-i\lambda_j t} = 0$ ,  $I(\lambda_j)$  is also the periodogram of  $\hat{u}_t$  at these frequencies, where  $\hat{u}_t = d_t - \bar{d}$ . Kernels  $k(j/M)$  in Equation (2) and  $K_M(\lambda_j)$  in Equation (7) are related, as  $K_M(\lambda) := (2\pi)^{-1} \sum_{|l| < T} k(l/M) e^{-il\lambda}$ , and the WCE in Equation (2) has frequency domain representation

$$\int_{-\pi}^{\pi} K_M(\lambda) I^*(\lambda) d\lambda,$$

where  $I^*(\lambda)$  is the periodogram of  $d_t - \bar{d}$ . Weighted covariance estimation and weighted periodogram estimation are therefore very similar, and this suggests for WPE an alternative theory analogue to fixed- $b$  for WCE.

<sup>1</sup>Where  $\lfloor \cdot \rfloor$  denotes the integer part of a number.



The WPE of the long-run variance using the Daniell kernel is

$$\hat{\gamma}_{\text{DAN}}^2 = 2\pi \frac{1}{m} \sum_{j=1}^m I(\lambda_j), \quad (8)$$

where  $m$  is a function of the bandwidth  $M$  (and, slightly abusing the notation, it is usually referred to as bandwidth itself). Regularity conditions, including  $m \rightarrow \infty$ , ensure that  $\hat{\gamma}_{\text{DAN}}^2$  is a consistent estimate of  $\sigma_T^2$ ; for fixed- $m$ , this is no longer the case but, using results from Hannan (1970), it is possible to show that

$$\sqrt{T} \frac{\bar{d}}{\hat{\gamma}_{\text{DAN}}} \rightarrow_d t_{2m} \quad (9)$$

(see also Brillinger, 1975, exercise 5.13.25; Müller, 2014; Sun, 2013). Regularity conditions in Müller (2014) include linearity, whereas in Sun (2013) these are milder, basically including just a central limit theorem (CLT) with constant variance  $\sigma^2$ . In Appendix A, we show that Equation (9) also holds under Assumptions 1-4. This warrants the applicability of fixed-smoothing asymptotics to a wide range of applications in forecast evaluation.

Monte Carlo simulations in Hualde and Iacone (2015) and Lazarus et al. (2018) show that fixed- $m$  asymptotics has the same size–power tradeoff documented for fixed- $b$  asymptotics: the smaller the value for  $m$ , the better the empirical size, but also the weaker the power.

## 5 | MONTE CARLO STUDY

In this section we revisit the size and power properties of the DM test in the framework of asymptotically nonvanishing estimation uncertainty of Giacomini and White (2006), focusing on the case of serial dependence.

We assume the following DGP:

$$y_t = \beta_t x_t + \varepsilon_t, \quad (10)$$

$$x_t = 0.5x_{t-1} + v_t, \quad (11)$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t, \quad (12)$$

$$v_t \sim \text{i.i.d. } N(0, 1), \quad u_t \sim \text{i.i.d. } N(0, \sigma^2), \quad \text{independent of each other}, \quad (13)$$

and we close the model by setting

$$\beta_{t+1} = \frac{1}{2}K_t + \frac{1}{2} \frac{x_t' \Omega(1) x_t}{K_t S_t^2} - \frac{\dot{\omega}(1)' x_t}{K_t S_t x_{t+1}^2}, \quad (14)$$

where

$$K_t = \frac{\sum_{j=t-R+1}^t \beta_j x_j^2}{\sum_{j=t-R+1}^t x_j^2}, \quad S_t = \sum_{j=t-R+1}^t x_j^2,$$

and  $\underline{x} = (x_{t-R+1}, \dots, x_t)'$ ,  $\underline{\varepsilon} = (\varepsilon_{t-R+1}, \dots, \varepsilon_t)'$ ,  $\Omega(\sigma) = E(\underline{\varepsilon}\underline{\varepsilon}')$ , and  $\dot{\omega}(\sigma) = (\omega_R(\sigma), \dots, \omega_1(\sigma))'$  with  $\omega_{t+1-j}(\sigma) = E(\varepsilon_j \varepsilon_{t+1-j})$ .

The DGP in Equations (10)–(14) is based on Giacomini and Rossi (2010), who, however, only considered the case of  $\rho = 0$ . As we are interested in analyzing the performance of standard and fixed-smoothing asymptotics when the autocorrelation of the loss differential increases, we augment the model of Giacomini and Rossi by adding dependence in Equation (12).

We give here an explanation of why the choice of  $\beta_{t+1}$  in Equation (14) is appropriate, and refer to Appendix B for the complete derivation. Following Giacomini and Rossi (2010), we use a quadratic loss function to compare one-step-ahead forecasts from two models: Model 1, where  $y_t = \beta x_t + v_t^{(1)}$ ; and Model 2, where  $y_t = v_t^{(2)}$ . The one-step-ahead forecasts of  $y_{t+1}$  implied by the two models are

$$\hat{y}_{t+1}^{(1)} = \hat{\beta}_{t,R} x_{t+1}, \quad (15)$$

$$\hat{y}_{t+1}^{(2)} = 0, \quad (16)$$

where  $\hat{\beta}_{t,R}$  is the in-sample parameter estimated based on a rolling window of size  $R$ , and where we assume for simplicity that  $x_{t+1}$  is known at time  $t$ .

The two models have equal predictive performance at the estimated parameter values in each point in time if

$$E \left\{ \left[ y_{t+1} - \hat{y}_{t+1}^{(1)} \left( \hat{\beta}_{t,R} \right) \right]^2 \right\} = E \left[ \left( y_{t+1} - \hat{y}_{t+1}^{(2)} \right)^2 \right]. \quad (17)$$

When  $\sigma^2 = 1$  in Equation (13), Giacomini and Rossi (2010) showed that, if  $\rho = 0$ , then Equation (17) holds for

$$\beta_{t+1} = \frac{1}{2}K_t + \frac{1}{2} \frac{1}{K_t S_t}. \quad (18)$$

For generic  $|\rho| < 1$  and  $\sigma^2 = 1$ , Equation (17) holds for Equation (14).<sup>2</sup> On the other hand, when  $\sigma^2 < 1$  in Equation (13), then, due to a reduction in the variance of the parameter estimate  $\hat{\beta}_{t,R}$ , Model 1 provides more accurate forecasts.

With this generalization of  $\beta_{t+1}$  to allow for nonzero  $\rho$ , we can analyze the size and power properties of the proposed tests for values of  $\rho$  that range from 0 to 0.8. To this end, we first generate time series for  $x_t$  and  $\varepsilon_t$  as in Equations (11)–(13). As in Giacomini and Rossi (2010), we initialize  $\beta_t$  with  $\beta_t = 0.05$  for  $t = 1, \dots, R$ . Then, using Equations (10) and (14) we generate  $n = R + T$  observations for  $y_t$ . To study size, we set  $\sigma^2 = 1$  in Equation (13); to study power we set  $\sigma^2 < 1$  in Equation (13). In all cases, we use 10,000 replications and the in-sample and out-of-sample sizes  $(R, T)$  equal to (50, 40) and (150, 120).

We consider three estimates of the long-run variance of the loss differential: the WCE using the DM estimate in Equation (3); the WCE using the Bartlett kernel in Equations (4)–(5); and the WPE using the Daniell kernel in Equation (8). We refer to these three estimates as WCE-DM, WCE-B, and WPE-D, respectively. For the WCE-DM in this Monte Carlo exercise, we set  $h = 1$  as in Giacomini and White (2006). For the WCE-B and the WPE-D estimates, we also need to specify a bandwidth. Newey and West (1994) showed that the optimal bandwidth, in minimal MSE sense, was proportional to  $M = \lfloor T^{1/3} \rfloor$ , so we use this bandwidth for the test with WCE-B and standard asymptotics. As for the tests with fixed-smoothing asymptotics, we use the bandwidths  $M = \lfloor T^{1/2} \rfloor$  for the WCE-B and  $m = \lfloor T^{1/3} \rfloor$  for the WPE-D, as we found in a preliminary Monte Carlo study that they give the best size–power combination when combined with fixed-smoothing asymptotics; see the Supporting Information Appendix for details.

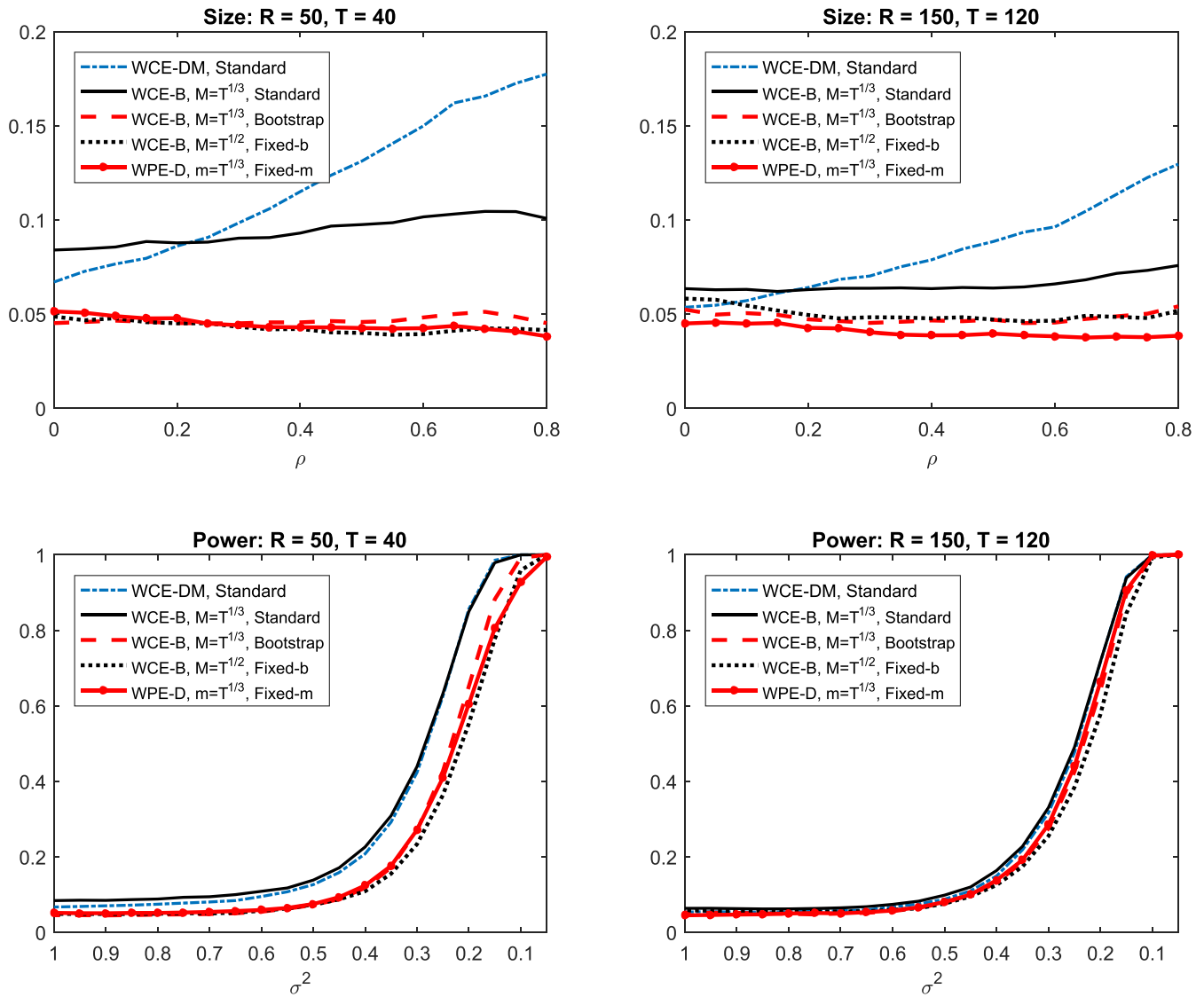
In the top plots of Figure 1, we report results of the Monte Carlo exercise with theoretical size set to 5%. We consider five different tests: two with standard normal limit distribution, one with bootstrap critical values, and two that use fixed-smoothing asymptotics. The test with bootstrap critical values uses a WCE-B estimate of the long-run variance and truncation  $M = \lfloor T^{1/3} \rfloor$ . Bootstrap critical values are computed using the overlapping stationary block-bootstrap of Politis and Romano (1994) with a circular scheme, as described in Appendix C.

Given that in this Monte Carlo  $h = 1$ , the test with WCE-DM estimate uses the sample variance to estimate the long-run variance, and for this reason it becomes seriously oversized as the degree of autocorrelation of the error increases. This problem is more serious in the Monte Carlo experiment with small in-sample and out-of-sample sizes, but it is substantial also for the rolling window of 150 observations and the out-of-sample size of 120, as shown by the 0.13 empirical size of the test for  $\rho = 0.8$ . This problem is partly addressed by the test that uses a WCE of the long-run variance with Bartlett kernel and truncation  $\lfloor T^{1/3} \rfloor$  and standard normal asymptotics. Still, also this test is oversized, especially for  $R = 50$  and  $T = 40$ . On the other hand, the tests that use bootstrap-critical values and fixed-smoothing asymptotics are correctly sized for all the rolling windows and out-of-sample sizes, and also for any degree of autocorrelation of the error.

To assess the power properties of the DM test with standard and fixed-smoothing asymptotics, we generate data under the alternative hypothesis that Model 1 provides more accurate forecasts by simulating the DGP in Equations (10)–(14) with  $\sigma^2$  in Equation (13) that decreases from its value of 1 when the null hypothesis is correct to 0.05. Since, in this case, we are interested in the power of the test, we set  $\rho = 0$  in all experiments. Results are reported in the bottom plots of Figure 1. As expected, the power of the tests increases towards 1 as  $\sigma^2$  decreases from 1 to 0.05. For small rolling windows and out-of-sample sizes, the tests that use standard asymptotics are oversized and have higher rejection frequencies for any  $\sigma^2$ . For larger rolling window and out-of-sample sizes, all tests are correctly sized and the plot shows that the power loss associated with the use of fixed-smoothing asymptotics is very small.

<sup>2</sup>Note that for  $\rho = 0$ , Equation (14) reduces to Equation (18).





**FIGURE 1** Size and power of the DM test. The figure displays empirical rejection frequencies for the DM test at 5% nominal size. In-sample and out-of-sample sizes  $(R, T)$  are equal to  $(50, 40)$  for the left-hand-side plots and  $(150, 120)$  for the right-hand-side plots. The top plots report the empirical size for different values of  $\rho$  in Equation (12). The bottom plots report the power of the tests for different values of  $\sigma^2$  in Equation (13). WCE-DM, Standard refers to the test that uses the sample variance to estimate the long-run variance and standard normal limit distribution. WCE-B,  $M=T^{1/3}$ , Standard refers to the test that uses a WCE of the long-run variance with Bartlett kernel with truncation  $\lfloor T^{1/3} \rfloor$  and standard normal limit distribution. WCE-B,  $M=T^{1/3}$ , Bootstrap refers to the test that uses a WCE of the long-run variance with Bartlett kernel with truncation  $\lfloor T^{1/3} \rfloor$  and bootstrap critical values, computed as detailed in Appendix C. WCE-B,  $M=T^{1/2}$ , Fixed-b refers to the test that uses WCE of the long-run variance with Bartlett kernel with  $M = \lfloor T^{1/2} \rfloor$  and fixed- $b$  asymptotics. WPE-D,  $m=T^{1/3}$ , Fixed-m refers to the test that uses WPE of the long-run variance with Daniell kernel with  $m = \lfloor T^{1/3} \rfloor$  and fixed- $m$  asymptotics [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



## 6 | EMPIRICAL ILLUSTRATION

To illustrate the usefulness of fixed-smoothing asymptotics for equal predictive accuracy tests, we evaluate the predictive accuracy of the Surveys of Professional Forecasters (SPF) and of the ECB Survey of Professional Forecasters (ECB SPF).

We perform the DM test, with WCE-DM, WCE-B, and WCE-D estimates of the long-run variance, using standard and fixed-smoothing asymptotics. To compute the WCE-DM, we use truncation lags equal to the forecast horizons as in Diebold and Mariano (1995). We use the bandwidths  $M = \lfloor T^{1/2} \rfloor$  for the WCE-B and  $m = \lfloor T^{1/3} \rfloor$  for the WPE-D, as in a preliminary Monte Carlo study they gave the best size–power combination; see the Supporting Information Appendix for details.

**TABLE 1** Real output growth: SPF versus random walk

Forecast horizon	0	1	2	3	4
<i>Evaluation period: 1987:Q1–2016:Q4, T = 120</i>					
WCE-DM	4.85	2.78	1.84	1.46	2.13
WCE-B, $M = \lfloor T^{1/2} \rfloor$	4.47**	2.64**	1.96*	1.76	2.24**
WPE-D, $m = \lfloor T^{1/3} \rfloor$	5.21**	2.63**	1.86*	1.68	2.08*
<i>Evaluation period: 1987:Q1–1996:Q4, T = 40</i>					
WCE-DM	3.61	2.28	2.51	1.51	3.35
WCE-B, $M = \lfloor T^{1/2} \rfloor$	3.92**	2.18*	2.52**	1.68	2.01*
WPE-D, $m = \lfloor T^{1/3} \rfloor$	4.01**	2.01*	2.72**	1.55	2.38*
<i>Evaluation period: 1997:Q1–2006:Q4, T = 40</i>					
WCE-DM	2.27	3.21	1.37	0.87	1.84
WCE-B, $M = \lfloor T^{1/2} \rfloor$	2.28*	3.01**	1.23	0.94	1.87
WPE-D, $m = \lfloor T^{1/3} \rfloor$	2.04*	2.71**	1.13	0.84	1.68
<i>Evaluation period: 2006:Q1–2016:Q4, T = 40</i>					
WCE-DM	2.79	1.63	1.14	0.99	1.12
WCE-B, $M = \lfloor T^{1/2} \rfloor$	2.19*	1.59	1.22	1.09	1.10
WPE-D, $m = \lfloor T^{1/3} \rfloor$	1.98*	1.43	1.08	0.94	0.98

*Note.* This table reports the predictive accuracy tests for the SPF forecasts of real GNP/GDP growth with respect to a random walk. GNP/GDP growth is defined as the annualized quarter-over-quarter growth rate of fixed-weighted real GNP in the surveys conducted before 1992:Q1, fixed-weighted real GDP in the surveys from 1992:Q1 to 1995:Q4, and chain-weighted real GDP in the surveys thereafter. Random walk predictions and realized values are computed accordingly. A positive entry means a higher average loss for the forecast made using the random walk. Asterisks \*\* and \* indicate two-sided significance at, respectively, the 5% and 10% level using fixed- $b$  asymptotics for WCE-B and fixed- $m$  asymptotics for WPE-D.  and  indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics (limit normality).

We use as benchmark a naive random walk—that is, a no-change benchmark using the vintages of data that were available to the public before the survey's deadline. To evaluate the forecasts, we use the first release as realized value and a quadratic loss function. A positive loss differential means a higher loss for the forecast made using the random walk; conversely, a negative entry means a higher loss for the forecast made using the SPF. In the next two subsections, we describe the survey data and the empirical results for, respectively, the SPF and the ECB SPF.

## 6.1 | Survey of Professional Forecasters

Data on the SPF is provided by the Federal Reserve Bank of Philadelphia and is available at a quarterly frequency. In particular, each quarter, the SPF asks panel members to make forecasts for a set of macroeconomic indicators for the current quarter and for the following four quarters. The deadline for panel members to submit their forecasts is the middle of the quarter. We focus on median responses for output growth, output inflation, the unemployment rate, and the 3-month Treasury bill for the period from 1987:Q1 until 2016:Q4 and consider four evaluation periods: the full sample and three 10-year subsamples—that is, from 1987:Q1 to 1996:Q4, from 1997:Q1 to 2006:Q4, and from 2007:Q1 to 2016:Q4. Breaking the sample into three subsamples is very important here as each subsample is associated with different historical features and different volatilities of the forecasts, with different impacts on the forecasting power of the SPF. The third subsample, for example, is associated with the Financial Crisis and forecasts from the SPF should fare comparatively better than the naive benchmark forecasts, as the SPF can readily take information into account, but this advantage may be offset by the higher volatility in the subsample, which may make it more difficult to obtain rejections of the null of equal forecasting power. However, as we have seen from the Monte Carlo exercise, the smaller dimension of the subsamples increases the risk of spurious evidence of significance using standard asymptotics, making the role of fixed-smoothing asymptotics very important.

In the SPF, the output price index is the implicit price deflator for gross national product (GNP) in surveys conducted prior to 1992:Q1, the implicit deflator for GDP in surveys from 1992:Q1 to 1995:Q4, and the chain-weighted price index in surveys conducted thereafter. In the same way, real output is defined as fixed-weighted real GNP in surveys conducted before 1992:Q1, fixed-weighted real GDP in surveys from 1992:Q1 to 1995:Q4, and chain-weighted real

**TABLE 2** GNP/GDP inflation: SPF versus random walk

Forecast horizon	0	1	2	3	4
<i>Evaluation period: 1987:Q1–2016:Q4, T = 120</i>					
WCE-DM	3.99	3.87	2.52	1.18	2.46
WCE-B, $M = \lfloor T^{1/2} \rfloor$	2.81**	3.68**	2.72**	1.02	2.54**
WPE-D, $m = \lfloor T^{1/3} \rfloor$	2.48**	3.53**	2.38**	0.86	2.25*
<i>Evaluation period: 1987:Q1–1996:Q4, T = 40</i>					
WCE-DM	1.31	2.42	1.22	0.04	1.31
WCE-B, $M = \lfloor T^{1/2} \rfloor$	1.36	2.37*	1.39	0.03	1.38
WPE-D, $m = \lfloor T^{1/3} \rfloor$	1.40	2.09*	1.55	0.03	1.17
<i>Evaluation period: 1997:Q1–2006:Q4, T = 40</i>					
WCE-DM	2.28	1.67	1.06	0.98	0.67
WCE-B, $M = \lfloor T^{1/2} \rfloor$	2.23*	1.66	0.99	1.01	0.78
WPE-D, $m = \lfloor T^{1/3} \rfloor$	2.25*	1.87	0.97	0.98	0.72
<i>Evaluation period: 2007:Q1–2016:Q4, T = 40</i>					
WCE-DM	3.66	3.17	2.10	0.97	3.13
WCE-B, $M = \lfloor T^{1/2} \rfloor$	2.75**	2.98**	2.29*	0.93	2.31*
WPE-D, $m = \lfloor T^{1/3} \rfloor$	2.47**	3.10**	1.91	0.86	2.25*

*Note.* this table reports the predictive accuracy tests for the SPF forecasts of GNP/GDP inflation with respect to a random walk. GNP/GDP inflation is defined as the implicit price deflator for GNP in surveys conducted prior to 1992:Q1, the implicit deflator for GDP in the surveys from 1992:Q1 to 1995:Q4, and the chain-weighted price index in the surveys thereafter. Random walk predictions and realized values are computed accordingly. A positive entry means a higher average loss for the forecast made using the random walk. Asterisks \*\* and \* indicate two-sided significance at, respectively, the 5% and 10% level using fixed- $b$  asymptotics for WCE-B and fixed- $m$  asymptotics for WPE-D.   and   indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics (limit normality).

GDP in surveys conducted thereafter. Real GNP/GDP growth and GNP/GDP inflation are constructed as the annualized quarter-over-quarter growth rates. For both variables, we define the corresponding benchmark forecasts and realized values accordingly, as in Stark (2010). Finally, the 3-month Treasury bill rate and the unemployment rate are expressed in levels.



Tables 1– 4 report the test statistics for the null hypothesis of equal predictive accuracy of the SPF forecasts for real GNP/GDP growth, GNP/GDP inflation, the unemployment rate, and the 3-month Treasury bill rates with respect to the random walk. In the tables, we use shades of gray to indicate two-sided significance using standard asymptotics (limit normality) and asterisks to indicate two-sided significance using fixed-smoothing asymptotics.

Results in Tables 1– 4 show that in the full sample the predictive ability of the SPF is stronger than that of the random walk for all the variables at short and medium horizons. The tables also indicate that for all the variables, with the exception of inflation, the first subsample is characterized by stronger predictive ability of the SPF with respect to the random walk than the other two subsamples. The results for the third subsample are most interesting, as we best see how standard asymptotics may lead to spurious rejections of the null hypothesis, and therefore to incorrect conclusions. To compare our results with the existing literature, Demetrescu et al. (2018) seem to provide the most noteworthy reference, as their sample was long enough to include the Financial Crisis and because they explicitly allowed for time-varying volatility. On the other hand, to test for time variation in an endogenous way, Demetrescu et al. (2018) employed a larger sample. Both the poor forecasting power in the second subsample and the weakness of the evidence of superior performance for the SPF in the third subsample are consistent with their findings.

We now discuss details of the various tests. Table 1 shows that the SPF forecasts for real GNP/GDP growth outperform the random walk on the full sample for short forecasting horizons. However, when looking at the three subsamples, the evidence of significant outperformance of the SPF is consistently supported by the tests with fixed- $b$  and fixed- $m$  asymptotics only for the nowcast. For the other horizons, the outperformance of the SPF sharply declined in the last subsample. As for GNP/GDP price inflation, Table 2 shows a much stronger predictive ability of the SPF, especially for short horizons and in the last subsample. Results in Table 3 indicate some predictive ability of the SPF forecasts for the unemployment rate, but the evidence is much weaker when using the proposed tests with fixed-smoothing asymptotics.

**TABLE 3** Unemployment rate: SPF versus random walk

Forecast horizon	0	1	2	3	4
<i>Evaluation period: 1987:Q1–2016:Q4, T = 120</i>					
WCE-DM	3.84	2.15	1.98	2.06	2.27
WCE-B, $M = \lfloor T^{1/2} \rfloor$	2.37**	1.99*	2.05*	2.20*	2.46**
WPE-D, $m = \lfloor T^{1/3} \rfloor$	2.14*	1.86*	1.88*	2.00*	2.20*
<i>Evaluation period: 1987:Q1–1996:Q4, T = 40</i>					
WCE-DM	3.47	1.72	1.66	2.07	2.52
WCE-B, $M = \lfloor T^{1/2} \rfloor$	3.13**	1.89	1.93	2.37*	2.57**
WPE-D, $m = \lfloor T^{1/3} \rfloor$	2.53**	1.61	1.58	1.89	2.01*
<i>Evaluation period: 1997:Q1–2006:Q4, T = 40</i>					
WCE-DM	2.17	1.75	1.47	1.23	1.15
WCE-B, $M = \lfloor T^{1/2} \rfloor$	2.23*	1.75	1.60	1.40	1.31
WPE-D, $m = \lfloor T^{1/3} \rfloor$	2.02*	1.51	1.36	1.17	1.09
<i>Evaluation period: 2007:Q1–2016:Q4, T = 40</i>					
WCE-DM	2.83	1.72	1.64	1.82	2.19
WCE-B, $M = \lfloor T^{1/2} \rfloor$	1.81	1.68	1.80	2.04*	2.42**
WPE-D, $m = \lfloor T^{1/3} \rfloor$	1.57	1.43	1.52	1.73	2.04*

*Note.* this table reports the predictive accuracy tests for the SPF forecasts of the unemployment rate with respect to a random walk. A positive entry means a higher average loss for the forecast made using the random walk. Asterisks \*\* and \* indicate two-sided significance at, respectively, the 5% and 10% level using fixed- $b$  asymptotics for WCE-B and fixed- $m$  asymptotics for WPE-D.  and  indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics (limit normality).



Finally, Table 4 provides strong evidence of superior predictive accuracy of the SPF forecasts for the 3-month Treasury bill rate with respect the random walk in the full sample. However, the predictive ability of the SPF for the 3-month Treasury bill rate declined sharply in the last two subsamples.

Comparing the application of standard asymptotics with fixed-smoothing asymptotics, we reject the null of equal predictive ability more frequently for the tests with standard asymptotics, especially for multistep forecasts in the subsamples. This is due to the fact that in the subsamples the tests are performed only on 40 observations and multistep forecast errors are more serially dependent, exacerbating the size distortions induced by standard asymptotics; see the Monte Carlo simulations in the Supporting Information Appendix. For example, Table 3 shows that for unemployment both the test with WCE-DM and test the with WCE-B and standard asymptotics reject at 10% significance level the null of equal predictive ability of the SPF and the random walk on the last subsample for almost all forecasting horizons. This could be interpreted as a clear indication of predictive ability of the SPF for the unemployment rate. However, the tests with fixed-smoothing asymptotics fail to reject the null of equal predictive ability for almost all forecasting horizons, especially when fixed- $m$  asymptotics is used, indicating that the SPF did not have much significant predictive ability for the unemployment rate in this period.

Finally, we note that results for the nowcasts (and, to a lesser extent, for longer horizons too) are affected by a common pitfall when using the WCE-DM, which is the presence of autocorrelation in the loss differential that is not accounted for by the WCE-DM estimate. Consider, for example, the nowcast for the unemployment rate in the last subsample, where  $T = 40$ . The WCE-DM test rejects the null of equal predictive ability of the SPF and the random walk with a test statistic equal to 2.83—much larger than the test statistics using the Bartlett kernel, which is 1.81. This is because the WCE-DM test statistic assumes optimality of the nowcasts and, as a consequence, uses only the sample variance to estimate the long-run variance. However, the sample first- and second-order autocorrelations of the loss differential are 0.665 and 0.298, and these are due to both autocorrelation in the forecast error of the SPF, which are respectively 0.258 and 0.214, and, more importantly, of the random walk, which are respectively 0.698 and 0.619. When we use the Bartlett kernel the estimates of the long-run variance are larger than the sample variance used to construct the WCE-DM statistic because the estimates of the long-run variance use also the first six autocovariances and these, given the degree of autocorrelation of the loss differential, are different from zero.

**TABLE 4** Three-month Treasury bill: SPF versus random walk

Forecast horizon	0	1	2	3	4
<i>Evaluation period: 1987:Q1–2016:Q4, T = 120</i>					
WCE-DM	5.05	3.79	3.16	2.72	1.76
WCE-B, $M = \lfloor T^{1/2} \rfloor$	3.84**	3.78**	3.54**	3.11**	1.96*
WPE-D, $m = \lfloor T^{1/3} \rfloor$	3.49**	3.58**	3.52**	3.27**	2.00*
<i>Evaluation period: 1987:Q1–1996:Q4, T = 40</i>					
WCE-DM	4.73	3.01	2.58	2.77	2.82
WCE-B, $M = \lfloor T^{1/2} \rfloor$	4.10**	3.50**	3.10**	2.94**	2.63**
WPE-D, $m = \lfloor T^{1/3} \rfloor$	3.17**	2.68**	2.56**	2.53**	2.10*
<i>Evaluation period: 1997:Q1–2006:Q4, T = 40</i>					
WCE-DM	3.02	2.46	2.10	1.69	1.07
WCE-B, $M = \lfloor T^{1/2} \rfloor$	2.18*	2.34*	2.28*	1.85	1.16
WPE-D, $m = \lfloor T^{1/3} \rfloor$	1.82	1.93	1.87	1.52	0.95
<i>Evaluation period: 2007:Q1–2016:Q4, T = 40</i>					
WCE-DM	1.68	1.19	1.14	0.77	-0.47
WCE-B, $M = \lfloor T^{1/2} \rfloor$	1.33	1.23	1.03	0.71	-0.46
WPE-D, $m = \lfloor T^{1/3} \rfloor$	1.17	1.09	0.92	0.65	-0.45

*Note.* this table reports the predictive accuracy tests for the SPF forecasts of the 3-month Treasury bill rate with respect to a random walk. A positive entry means a higher average loss for the forecast made using the random walk. Asterisks \*\* and \* indicate two-sided significance at, respectively, the 5% and 10% level using fixed- $b$  asymptotics for WCE-B and fixed- $m$  asymptotics for WPE-D.  and  indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics (limit normality).

## 6.2 | ECB Survey of Professional Forecasters

Data on the ECB SPF have been provided by European Central Bank since 1999. The survey is performed quarterly and includes expectations for some of the key macroeconomic variables for the euro area. Between 1999:Q1 and 2001:Q3, the survey was conducted in the middle month of each quarter—that is, in February, May, August, and November. Since 2001:Q4, the survey has been shifted to the first month of the quarter—that is, in January, April, July, and October. The questionnaire is sent to the panelists just after the release of the Harmonized Index of Consumer Prices (HICP)—that is, in the third week of the month before the survey—and the forecasts are collected in the second half of the first month of each quarter. For more details, Bowles et al. (2007, 2010), and Garcia (2003).

We focus on mean responses about the year-on-year GDP growth and year-on-year HICP inflation at the rolling horizons of 1, 2, and 5 years.<sup>3</sup> In the ECB SPF, the rolling horizons are set to 1, 2, and 5 years ahead of the latest period for which the variable in question is observed when the survey is conducted and not ahead of the survey date. For example, in the survey of the first quarter of 2008 (sent out after the official release of the December 2007 figure for HICP and of the 2007:Q3 figure for GDP), the questionnaire asked for the expected year-on-year inflation rate in December 2008 and the year-on-year GDP growth in 2008:Q3. Observations for the three forecasting horizons are only available from 2006:Q1 to 2016:Q4, for a total of 44 observations. With such a small sample size, standard tests of equal predictive ability suffer from large size distortions but fixed-smoothing asymptotics can still provide reliable inference.

Since data on GDP and HICP are subject to revisions, following Conflitti et al. (2015), we use the euro area real-time database (see Giannone et al., 2012) to match the survey data with the information that was available to the forecasters at the time that they submitted their projections.



Table 5 reports the test statistics presented in Section 4 for the null hypothesis of equal predictive accuracy of the ECB SPF forecasts for year-on-year GDP growth and year-on-year HICP inflation with respect to the random walk. As in Tables 1– 4, we use asterisks to indicate two-sided significance using fixed-smoothing asymptotics, and shades of gray to indicate two-sided significance using standard asymptotics and limit normality.

Results in Table 5 indicate that the tests with standard asymptotics reject the null of equal predictive ability of the ECB SPF with respect to random walk for year-on-year GDP growth at short and medium horizons, and also for year-on-year

<sup>3</sup>In the ECB SPF the 5-year horizon forecast refers to the full year. However, we take this horizon as the long-run forecast and treat it as a rolling window forecast.

TABLE 5 ECB Survey of Professional Forecasters

Forecast horizon (years)	GDP growth			HICP inflation		
	1	2	5	1	2	5
WCE-DM	2.09	1.80	1.04	1.28	1.75	1.04
WCE-B, $M = \lfloor T^{1/2} \rfloor$	1.84	2.04*	1.08	1.18	1.79	0.89
WPE-D, $m = \lfloor T^{1/3} \rfloor$	1.54	1.81	0.89	0.97	1.48	0.71

*Note.* this table reports the predictive accuracy tests for the ECB Survey of Professional Forecasts for the year-on-year GDP growth (left panel) and year-on-year HICP inflation (right panel) with respect to a random walk. The evaluation period is 2006:Q1 to 2016:Q4, for a total of 44 observations. A positive entry means a higher average loss for the forecast made using the random walk. Asterisks \*\* and \* indicate two-sided significance at, respectively, the 5% and 10% level using fixed- $b$  asymptotics for WCE-B and fixed- $m$  asymptotics for WPE-D.  and  indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics (limit normality).

HICP inflation at medium horizon. However, these results are partially spurious and demonstrate the risks of using standard asymptotics in a small sample. Indeed, when using fixed-smoothing asymptotics, we only find limited evidence of superior predictive ability of the ECB SPF with respect to the random walk and only for the GDP growth at medium horizon.

## 7 | CONCLUSION

We propose fixed-smoothing asymptotics to overcome the small-sample size distortions of standard tests for predictive accuracy. As an illustrative example, and to facilitate comparison with other works, we apply fixed-smoothing asymptotics to reassess the predictive accuracy of the SPF. We also include an application to the ECB SPF, which has a short time series dimension and thus makes our approach very convenient.

We focus on applying fixed-smoothing asymptotics to the Diebold and Mariano (1995) test. However, this approach is of broader applicability in the forecasting evaluation literature: Demetrescu et al. (2018) applied the framework to the forecast breakdown tests of Giacomini et al. (2009), and Coroneo et al. (2019) applied fixed-smoothing asymptotics to density forecast evaluation. Future work will include applying fixed-smoothing asymptotics to forecast rationality tests (see Granger & Newbold, 1986 and Diebold and Lopez ; 1996).

## ACKNOWLEDGMENTS

The authors thank the Editor, Barbara Rossi, and two anonymous referees for their very constructive and insightful comments, which helped improve the paper substantially. For helpful suggestions we are also grateful to Jia Chen, Valentina Corradi, David De Antonio Liedo, Bruce Hansen, Michael McCracken, and seminar participants at the University of York, University of Nottingham, University of Bologna, University of Hannover, the 2015 International Conference on Computational and Financial Econometrics, the 2016 Royal Economic Society Annual Conference, the 2016 International Symposium of Forecasting, the Applied Time Series Econometrics Workshop (Federal Reserve Bank of St. Louis), the workshop on Time Series Econometrics (Barcelona Graduate School of Economics), the Vienna Workshop on Economic Forecasting, 24th Conference in Computing in Economics and Finance, and the 8th Italian Congress of Econometrics and Empirical Economics. The support of the ESRC grant ES/K001345/1 is gratefully acknowledged.

## OPEN RESEARCH BADGES



This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at [<http://qed.econ.queensu.ca/jae/datasets/coroneo002/>].



## REFERENCES

- Bowles, C., Friz, R., Genre, V., Kenny, G., Meyler, A., & Rautanen, T. (2007). The ECB survey of professional forecasters (SPF): a review after eight years' experience. (*ECB Occasional Paper Series No. 59*): Frankfurt, Germany: European Central Bank.
- Bowles, C., Friz, R., Genre, V., Kenny, G., Meyler, A., & Rautanen, T. (2010). An evaluation of the growth and unemployment forecasts in the ECB Survey of Professional Forecasters. *Journal of Business Cycle Measurement and Analysis*, 2(4), 1–28.
- Brillinger, D. R. (1975). Time series data analysis and theory.
- Clark, T. E. (1999). Finite-sample properties of tests for equal forecast accuracy. *Journal of Forecasting*, 18(7), 489–504.
- Clark, T. E., & McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105(1), 85–110.
- Clark, T. E., & McCracken, M. W. (2013). Advances in forecast evaluation, *Handbook of economic forecasting* (pp. 1107–1201), Vol. 2. Amsterdam, Netherlands: Elsevier.
- Conflitti, C., De Mol, C., & Giannone, D. (2015). Optimal combination of survey forecasts. *International Journal of Forecasting*, 31(4), 1096–1103.
- Coroneo, L., Iacone, F., & Profumo, F. (2019). A real-time density forecast evaluation of the ECB Survey of Professional Forecasters. (*Discussion Papers 19/14*): York, UK: Department of Economics, University of York.
- Demetrescu, M., Hanck, C., & Kruse, R. (2018). Robust inference under time-varying volatility: A real-time evaluation of professional forecasters. (*Technical report*).
- Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. *Journal of Business and Economic Statistics*, 33(1), 1–1.
- Diebold, F. X., & Lopez, J. A. (1996). Forecast evaluation and combination, *Handbook of statistics* (pp. 241–268), Vol. 14. Amsterdam, Netherlands: North-Holland.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 20(1), 253–263.
- Garcia, J. A. (2003). An introduction to the ECB's Survey of Professional Forecasters. (*ECB Occasional Paper Series No. 8*): Frankfurt, Germany: European Central Bank.
- Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25(4), 595–620.
- Giacomini, R., Rossi, B., & Review of Economic Studies (2009). Detecting and predicting forecast breakdowns, 76(2), 669–705.
- Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545–1578.
- Giannone, D., Henry, J., Lalik, M., & Modugno, M. (2012). An area-wide real-time database for the euro area. *Review of Economics and Statistics*, 94(4), 1000–1013.
- Gonçalves, S., & Vogelsang, T. J. (2011). Block bootstrap hac robust tests: The sophistication of the naive bootstrap. *Econometric Theory*, 27(4), 745–791.
- Granger, C. W. J., & Newbold, P. (1986). *Forecasting economic time series* (2nd ed.). Orlando, FL: Academic Press.
- Hannan, E. J. (1970). *Multiple time series*. Chichester, UK: Wiley.
- Harvey, D. I., Leybourne, S. J., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13(2), 281–291.
- Harvey, D. I., Leybourne, S. J., & Whitehouse, E. J. (2017). Forecast evaluation tests and negative long-run variance estimates in small samples. *International Journal of Forecasting*, 33(4), 833–847.
- Hualde, J., & Iacone, F. (2015). Autocorrelation robust inference using the Daniell kernel with fixed bandwidth. (*Technical report*): York, UK: Department of Economics, University of York.
- Hualde, J., & Iacone, F. (2017). Fixed bandwidth asymptotics for the studentized mean of fractionally integrated processes. *Economics Letters*, 150, 39–43.
- Kiefer, N. M., & Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity–autocorrelation robust tests. *Econometric Theory*, 21(6), 1130–1164.
- Lazarus, E., Lewis, D. J., Stock, J. H., & Watson, M. W. (2018). Har inference: recommendations for practice. *Journal of Business and Economic Statistics*, 36(4), 541–559.
- Li, J., & Patton, A. J. (2018). Asymptotic inference about predictive accuracy using high frequency data. *Journal of Econometrics*, 203(2), 223–240.
- Müller, U. K. (2014). HAC corrections for strongly autocorrelated time series. *Journal of Business and Economic Statistics*, 32(3), 311–322.
- Newey, W. K., & West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61(4), 631–653.
- Patton, A. J. (2015). Comment. *Journal of Business and Economic Statistics*, 33(1), 22–24.
- Phillips, P. C. (2005). HAC estimation by automated regression. *Econometric Theory*, 21(1), 116–142.
- Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428), 1303–1313.
- Stark, T. (2010). Realistic evaluation of real-time forecasts in the survey of professional forecasters, *Research Rap, Special Report 1*.
- Sun, Y. (2013). A heteroskedasticity and autocorrelation robust  $F$  test using an orthonormal series variance estimator. *Econometrics Journal*, 16(1), 1–26.
- Sun, Y. (2014a). Fixed-smoothing asymptotics in a two-step generalized method of moments framework. *Econometrica*, 82(6), 2327–2370.
- Sun, Y. (2014b). Let's fix it: Fixed- $b$  asymptotics versus small- $b$  asymptotics in heteroskedasticity and autocorrelation robust inference. *Journal of Econometrics*, 178, 659–677.
- Sun, Y., Phillips, P. C., & Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing. *Econometrica*, 76(1), 175–194.
- West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64(5), 1067–1084.

West, K. D. (2006). Forecast evaluation, *Handbook of economic forecasting*, 1, 99–134.

Wooldridge, J. M., & White, H. (1988). Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory*, 4(2), 210–230.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Coroneo L, Iacone F. Comparing predictive accuracy in small samples using fixed-smoothing asymptotics. *J Appl Econ*. 2020;1–19. <https://doi.org/10.1002/jae.2756>

## APPENDIX A: FIXED-SMOOTHING ASYMPTOTICS FOR FORECAST EVALUATION

- Fixed- $m$ : In this case, under regularity conditions, such as strict stationarity of  $d_t$ , existence of the second moment and  $0 < \sum_{j=-\infty}^{\infty} \text{cov}(d_t, d_{t+j}) < \infty$ , then Equation (9) holds (see, e.g., Brillinger, 1975).

To see that it also holds under Assumptions 1-4, denote as  $W(r)$  a standard Brownian motion, and consider the functional central limit theorem (FCLT)  $\frac{1}{\sqrt{T}} \frac{1}{\sigma_T} \sum_{t=1}^{\lfloor rT \rfloor} d_t \Rightarrow W(r)$ , where  $\lfloor \cdot \rfloor$  denotes the integer part of a number, and  $r \in [0, 1]$ . This FCLT can be established for  $d_t$  from Assumptions 1-4: in particular, Giacomini and White (2006, p. 1575) noted that conditions such as Assumptions 1-4 satisfy the conditions of corollary 3.1 of Wooldridge and White (1988) FCLT and CLT for mixing processes.

Denote, as in Hualde and Iacone (2017),  $\hat{W}(r) = W(r) - rW(1)$  and

$$Q(j) = \left\{ \left( 2\pi j \int_0^1 \sin(2\pi jr) \hat{W}(r) dr \right)^2 + \left( 2\pi j \int_0^1 \cos(2\pi jr) \hat{W}(r) dr \right)^2 \right\}.$$

Recall that  $\hat{u}_t = d_t - \bar{d}$ . Using summation by parts as in Hualde and Iacone (2017), rewrite the periodogram as

$$I(\lambda_j) = \frac{1}{2\pi} \left\{ \sum_{t=1}^{T-1} \left[ -\frac{2\pi j}{T} \sin\left(2\pi j \frac{t}{T}\right) + O(\lambda_j^2) \right] \frac{1}{T^{1/2}} \sum_{s=1}^t \hat{u}_s \right\}^2 + \frac{1}{2\pi} \left\{ \sum_{t=1}^{T-1} \left[ \frac{2\pi j}{T} \cos\left(2\pi j \frac{t}{T}\right) + O(\lambda_j^2) \right] \frac{1}{T^{1/2}} \sum_{s=1}^t \hat{u}_s \right\}^2,$$

then, by the FCLT and the continuous mapping theorem, as  $T \rightarrow \infty$ ,

$$\frac{2\pi}{\sigma_T^2} I(\lambda_j) \Rightarrow Q(j).$$

Combining these results, another application of the continuous mapping theorem yields

$$\sqrt{T} \frac{\bar{d}}{\sqrt{2\pi \frac{1}{m} \sum_{j=1}^m I(\lambda_j)}} = \frac{\sqrt{T} \frac{1}{\sigma_T} \bar{d}}{\sqrt{2\pi \frac{1}{\sigma_T^2} \frac{1}{m} \sum_{j=1}^m I(\lambda_j)}} \rightarrow_d \frac{W(1)}{\sqrt{\frac{1}{m} \sum_{j=1}^m Q(j)}},$$

which is indeed the same limit as in Hualde and Iacone (2017), and note that this is a  $t_{2m}$  distributed variable.

- Fixed- $b$ : The proof for fixed- $b$  asymptotics uses the same FCLT, to establish the limit in Kiefer and Vogelsang (2005).

## APPENDIX B: THE UNCONDITIONAL PREDICTIVE ABILITY TEST IN PRESENCE OF SERIAL CORRELATION

Consider the DGP in Equations (10)–(13), and the two competing forecasts in Equations (15)–(16) where

$$\hat{\beta}_{t,R} = \frac{\sum_{j=t-R+1}^t y_j x_j}{\sum_{j=t-R+1}^t x_j^2} = \frac{\sum_{j=t-R+1}^t \beta_j x_j^2}{\sum_{j=t-R+1}^t x_j^2} + \frac{\sum_{j=t-R+1}^t \varepsilon_j x_j}{\sum_{j=t-R+1}^t x_j^2}.$$

Note that  $\beta_t$  in Equation (10) changes at any point  $t$ , but the estimate is made with OLS, which is designed for  $\beta$  constant over the period  $t - R + 1, \dots, t$ . However, the estimate  $\hat{\beta}_{t,R}$  changes at any point in  $t$  because of its recursive structure.

To derive the values of  $\beta_t$  that ensure that the two models have equal predictive performance at the estimated parameters values, as prescribed in Equation (17), we first revise the derivation for the case of no serial dependence, as in Giacomini and Rossi (2010).

The forecast error of Model 1 is given by

$$y_{t+1} - \hat{y}_{t+1}^{(1)} = \beta_{t+1} x_{t+1} + \varepsilon_{t+1} - \hat{\beta}_{t,R} x_{t+1} = (\beta_{t+1} - \hat{\beta}_{t,R}) x_{t+1} + \varepsilon_{t+1},$$

with

$$\left( y_{t+1} - \hat{y}_{t+1}^{(1)} \right)^2 = \left( \beta_{t+1} - \frac{\sum_{j=t-R+1}^t \beta_j x_j^2}{\sum_{j=t-R+1}^t x_j^2} - \frac{\sum_{j=t-R+1}^t \varepsilon_j x_j}{\sum_{j=t-R+1}^t x_j^2} \right)^2 x_{t+1}^2 + \varepsilon_{t+1}^2 + 2 \left( \beta_{t+1} - \frac{\sum_{j=t-R+1}^t \beta_j x_j^2}{\sum_{j=t-R+1}^t x_j^2} - \frac{\sum_{j=t-R+1}^t \varepsilon_j x_j}{\sum_{j=t-R+1}^t x_j^2} \right) x_{t+1} \varepsilon_{t+1}$$

and note that  $\varepsilon_{t+1}$  and  $\varepsilon_j$  for  $j = t - R + 1, \dots, t$  has  $E(\varepsilon_{t+1}, \varepsilon_j) = 0$ , so the expected value of the third term is 0.

As for the first term,

$$\left( \beta_{t+1} - \frac{\sum_{j=t-R+1}^t \beta_j x_j^2}{\sum_{j=t-R+1}^t x_j^2} - \frac{\sum_{j=t-R+1}^t \varepsilon_j x_j}{\sum_{j=t-R+1}^t x_j^2} \right)^2 = \left( \beta_{t+1} - \frac{\sum_{j=t-R+1}^t \beta_j x_j^2}{\sum_{j=t-R+1}^t x_j^2} \right)^2 + \left( \frac{\sum_{j=t-R+1}^t \varepsilon_j x_j}{\sum_{j=t-R+1}^t x_j^2} \right)^2 - 2 \left( \beta_{t+1} - \frac{\sum_{j=t-R+1}^t \beta_j x_j^2}{\sum_{j=t-R+1}^t x_j^2} \right) \left( \frac{\sum_{j=t-R+1}^t \varepsilon_j x_j}{\sum_{j=t-R+1}^t x_j^2} \right) \quad (\text{B.1})$$

and again the expected value of the third term is 0. The first term is nonrandom, while the second one is

$$\frac{\left( \sum_{j=t-R+1}^t \varepsilon_j x_j \right)^2}{\left( \sum_{j=t-R+1}^t x_j^2 \right)^2} = \frac{\left( \sum_{j,k=t-R+1}^t \varepsilon_j x_j \varepsilon_k x_k \right)}{\left( \sum_{j=t-R+1}^t x_j^2 \right)^2}$$

and, as  $\varepsilon_j$  is independently distributed,  $E\left(\sum_{j,k=t-R+1}^t \varepsilon_j x_j \varepsilon_k x_k\right) = \sigma^2 \sum_{j=t-R+1}^t x_j^2$ , so the expectation of the second term is

$\left( \sum_{j=t-R+1}^t x_j^2 \right)^{-1} \sigma^2$ . Thus

$$E\left( y_{t+1} - \hat{y}_{t+1}^{(1)} \right)^2 = \left\{ \left( \beta_{t+1} - \frac{\sum_{j=t-R+1}^t \beta_j x_j^2}{\sum_{j=t-R+1}^t x_j^2} \right)^2 + \frac{\sigma^2}{\sum_{j=t-R+1}^t x_j^2} \right\} x_{t+1}^2 + \sigma^2.$$

For the other forecast,

$$\begin{aligned} E\left(y_{t+1} - \hat{y}_{t+1}^{(2)}\right)^2 &= E(\beta_{t+1}x_{t+1} + \varepsilon_{t+1})^2 \\ &= E(\beta_{t+1}^2x_{t+1}^2 + \varepsilon_{t+1}^2 + 2\beta_{t+1}x_{t+1}\varepsilon_{t+1}) \\ &= \beta_{t+1}^2x_{t+1}^2 + \sigma^2. \end{aligned}$$

Imposing  $E\left(y_{t+1} - \hat{y}_{t+1}^{(1)}\right)^2 = E\left(y_{t+1} - \hat{y}_{t+1}^{(2)}\right)^2$ , as in Equation (17), we obtain

$$\left\{(\beta_{t+1} - K_t)^2 + \frac{\sigma^2}{S_t}\right\}x_{t+1}^2 + \sigma^2 = \beta_{t+1}^2x_{t+1}^2 + \sigma^2,$$

with

$$K_t = \frac{\sum_{j=t-R+1}^t \beta_j x_j^2}{\sum_{j=t-R+1}^t x_j^2}, \quad S_t = \sum_{j=t-R+1}^t x_j^2.$$

Solving for  $\beta_{t+1}$  we get

$$\beta_{t+1} = \frac{1}{2}K_t + \frac{1}{2}\frac{\sigma^2}{K_t S_t},$$

as indeed in Giacomini and Rossi (2010); setting  $\sigma^2 = 1$ , we obtain the value given in Equation (18).

For the case in which  $\varepsilon_t$  is a dependent process, consider again the second term in Equation (B.1),  $E\left(\sum_{j,k=t-R+1}^t \varepsilon_j x_j \varepsilon_k x_k\right)$ .

Letting  $\underline{\varepsilon} = (\varepsilon_{t-R+1}, \dots, \varepsilon_t)'$ ,  $\underline{x} = (x_{t-R+1}, \dots, x_t)'$ , then

$$\sum_{j,k=t-R+1}^t \varepsilon_j x_j \varepsilon_k x_k = (\underline{\varepsilon}' \underline{x})^2 = \underline{\varepsilon}' \underline{x} \underline{\varepsilon}' \underline{x} = \underline{x}' \underline{\varepsilon} \underline{\varepsilon}' \underline{x},$$

because  $\underline{\varepsilon}' \underline{x} = \underline{x}' \underline{\varepsilon}$  as these are scalars. So, letting  $\Omega = \Omega(\sigma) = E(\underline{\varepsilon} \underline{\varepsilon}')$

$$E\left(\sum_{j,k=t-R+1}^t \varepsilon_j x_j \varepsilon_k x_k\right) = \underline{x}' \Omega \underline{x}.$$

Also, letting  $\omega_{t+1-j} = \omega_{t+1-j}(\sigma) = E(\varepsilon_j \varepsilon_{t+1})$  then

$$E\left(\sum_{j=t-R+1}^t \varepsilon_j x_j x_{t+1} \varepsilon_{t+1}\right) = \sum_{j=t-R+1}^t \omega_{t+1-j} x_j x_{t+1}.$$

Then

$$\begin{aligned}
 E\left(y_{t+1} - \hat{y}_{t+1}^{(1)}\right)^2 &= E\left[\beta_{t+1} - \frac{\sum_{j=t-R+1}^t \beta_j x_j^2}{\sum_{j=t-R+1}^t x_j^2} - \frac{\sum_{j=t-R+1}^t \varepsilon_j x_j}{\sum_{j=t-R+1}^t x_j^2}\right]^2 x_{t+1}^2 + E\varepsilon_{t+1}^2 + 2E\left[\beta_{t+1} - \frac{\sum_{j=t-R+1}^t \beta_j x_j^2}{\sum_{j=t-R+1}^t x_j^2} - \frac{\sum_{j=t-R+1}^t \varepsilon_j x_j}{\sum_{j=t-R+1}^t x_j^2}\right] x_{t+1} \varepsilon_{t+1} \\
 &= (\beta_{t+1} - K_t)^2 x_{t+1}^2 + \frac{E\left(\sum_{j=t-R+1}^t \varepsilon_j x_j\right)^2}{\left(\sum_{j=t-R+1}^t x_j^2\right)^2} x_{t+1}^2 - 2\left[\beta_{t+1} - \frac{\sum_{j=t-R+1}^t \beta_j x_j^2}{\sum_{j=t-R+1}^t x_j^2}\right] \left[\frac{E\sum_{j=t-R+1}^t \varepsilon_j x_j}{\sum_{j=t-R+1}^t x_j^2}\right] x_{t+1}^2 + \omega_0 + \\
 &\quad - 2 \frac{E\left(\sum_{j=t-R+1}^t \varepsilon_j x_j\right) \varepsilon_{t+1} x_{t+1}}{\sum_{j=t-R+1}^t x_j^2} \\
 &= (\beta_{t+1} - K_t)^2 x_{t+1}^2 + \frac{x' \Omega x}{S_t^2} x_{t+1}^2 + \omega_0 - 2 \frac{\dot{\omega}' x x_{t+1}}{S_t},
 \end{aligned}$$

where  $\dot{\omega}' x x_{t+1} = \sum_{j=t-R+1}^t \omega_{t+1-j} x_j x_{t+1}$ .

For the other forecast,

$$E\left(y_{t+1} - \hat{y}_{t+1}^{(1)}\right)^2 = E\left(\beta_{t+1}^2 x_{t+1}^2 + \varepsilon_{t+1}^2 + 2\beta_{t+1} x_{t+1} \varepsilon_{t+1}\right) = \beta_{t+1}^2 x_{t+1}^2 + \omega_0^2$$

and, again imposing Equation (17),

$$(\beta_{t+1} - K_t)^2 x_{t+1}^2 + \frac{x' \Omega x}{S_t^2} x_{t+1}^2 + \omega_0 - 2 \frac{\dot{\omega}' x x_{t+1}}{S_t} = \beta_{t+1}^2 x_{t+1}^2 + \omega_0$$

and, solving for  $\beta_{t+1}$ , we get

$$\beta_{t+1} = \frac{1}{2} K_t + \frac{1}{2} \frac{x' \Omega x}{K_t S_t^2} - \frac{\dot{\omega}' x x_{t+1}}{K_t S_t x_{t+1}^2}.$$

## APPENDIX C: BOOTSTRAP DETAILS

Given the loss differential  $d_t$  and the test statistic

$$t = \sqrt{T} (\bar{d} / \hat{\sigma}), \quad (C1)$$

where  $\hat{\sigma}$  is the estimate of the long-run variance computed using the WCE-B or the WPE-D, for each bootstrap replication  $r$ , we generate bootstrapped loss differentials  $d_t^{(r)}$  using the overlapping stationary block-bootstrap of Politis and Romano (1994) with a circular scheme. In particular, in each bootstrap replication  $r$ , we perform the following steps:

1. Draw block sizes  $L_1, L_2, \dots$  from a discrete uniform distribution with support on  $\{1, \dots, 2 \lfloor T^{1/4} \rfloor\}$ .
2. Draw random initial indices  $I_1, I_2, \dots$  from a discrete uniform distribution with support on  $\{1, \dots, T\}$ .
3. Construct the series of bootstrapped loss differential  $d_t^{(r)}$  using the first  $T$  elements of  $(\tilde{d}_{I_1}, \dots, \tilde{d}_{I_1+L_1-1}, \tilde{d}_{I_2}, \dots, \tilde{d}_{I_2+L_2-1}, \dots)$ , where  $\tilde{d}_t$  is the collated loss differential; that is,  $\tilde{d}_t \equiv (d_1, \dots, d_T, d_1, \dots, d_T)$ .
4. Construct the bootstrapped test statistic as

$$t^{(r)} = \sqrt{T} ((\bar{d}^{(r)} - \bar{d}) / \hat{\sigma}^{(r)}),$$

where  $\bar{d}^{(r)}$  is the sample mean of  $d_t^{(r)}$ , and  $\hat{\gamma}^{(r)}$  is the estimate of its long-run variance constructed using the same formula as in Equation (C1).

We perform 10,000 bootstrap replications and use the 95% quantile of the bootstrap distribution of the test statistic,  $(|t^{(1)}|, \dots, |t^{(10,000)}|)$ , as critical value  $cv$ . We then reject the null of equal predictive ability if  $|t| > cv$ . Note that this is the naive bootstrap also performed by Kiefer and Vogelsang (2005) and Gonçalves and Vogelsang (2011) for the test with the WCE-B estimate of the long-run variance.